確率モデルに基づく2値分類から多値分類へのデコード(情報物理学の数学的構造)

AUTHOR(S):

石井, 信

CITATION:

石井, 信. 確率モデルに基づく2値分類から多値分類へのデコード(情報物理学の数学的構造). 数理解析研究所講究録 2007, 1532: 11-18

ISSUE DATE:

2007-02

URL:

http://hdl.handle.net/2433/58961

数理解析研究所講究録
1532巻 2007年 11-18

確率モデルに基づく2値分類から多値分類へのデコード

奈良先端科学技術大学院大学 情報科学研究科　石井 信 [1]

1 確率モデルに基づく多値分類

多値 (特に M 値とする) 分類問題とは、各データ点 x をあらかじめ用意された複数のクラス C ($|C| = M$) のどれか一つに分類する問題である。x を条件としたクラス所属の事後確率 $P(i|x)$ ($i \in C$) が既知である場合、事後確率を判別関数として、データ点 $x^{(n)}$ ごとに事後確率を最大にするクラス $c(x^{(n)}) = \arg\max_i P(i|x^{(n)})$ を選択すれば、判別正解率が最大になる。これをベイズ最適決定と呼び、分類結果をベイズ最適決定に近付けることが多値分類の目的である。クラスの事後確率が未知の場合、それを直接近似しようする方法として、正規混合分布などのパラメトリックな生成モデルに基く方法や、カーネル密度推定や K 最近傍法などのノンパラメトリックな方法が用いられる。

一方で、2値 ($M = 2$) 分類問題に限れば、2つのクラス間の決定境界のマージンを制御することによって、クラス分類の汎化性能を高めることができる。Support vector machine (SVM), AdaBoost など、次元が高い場合やベイズ最適決定境界が複雑な形状をしている場合にも高い判別性能が得られることが多いため、広く用いられている。しかしこれらは2値分類問題に特化した手法であるため、多値分類問題に応用するためには、多値分類問題を複数の2値分類問題に分割して、後に統合するというヒューリスティクスが用いられてきた。M クラスの分類問題を、1クラスとその残りを分類する問題に分ける方法 (one-versus-the rest, 1R)、あるいは、一対一の分類問題に分ける方法 (one-versus-one, 11) が、計算量が比較的少なくかつ簡便なものとして知られている。また、複数の2値分類器の各々に対して、2つのクラスへの分類結果を正 (+) と負 (−) への符号割り当てとみなした符号行列を構成しておき、2値分類器の判別結果を並べた符号語から、誤り訂正符号方式に基いて、多値分類結果へ復号化させる枠組みも提案されている [4]。さらに、[1] や [9] では、符号行列の各要素が $(+, -, 0)$ の3値になることで、2つの任意のクラス集合を分類対象とする場合を扱っている。

本報告では、任意の2値分類器の出力結果からなる集合から多値分類器の出力に復号化するための統計的手法を紹介する。2値分類器の出力は多くの場合、まずはアナログ値である判別関数 $f(x)$

[1] E-mail: ishii@is.naist.jp

として与えられるが、復号化を考える際に、(a) 2値化関数 $\mathrm{sig}(f(x))$ により、+, − に2値化して扱う、(b) 判別関数値のまま直接扱う、もしくは (c) 何らかの手法 (例えば logistic regression) による変換により確率値にして扱うか、の選択肢がある。場合 (a) では、単純な投票法による復号過程が考えられ、これは符号語と符号行列との間で Hamming 距離を最小にするような復号化を行う Hamming decoding [4] と等価になる。Allwein ら [1] はまた、場合 (b) として、Hamming 距離でなく判別関数の値に基づく距離も提案し loss-based decoding と呼んだ。場合 (c) について、Hastie ら [5] は、11 の符号表に限った、すなわち、2 値分類器として 11 のみを用いた場合について、逐次最適化アルゴリズムによる復号法を提案した。Zadrozny [9] は、これを任意の符号表に対して使用できるように拡張するとともに、Hastie ら [5] の手法が、11 の符号表に対しては、loss-based decoding と等価であることを示した。

以上の背景に基づき、本報告では、複数の 2 値分類器の組み合わせにより最適な多値分類器を構成するための、復号化と符号化に関する統計的手法について述べる。まず、与えられた 2 値分類結果と真の多値分類結果から規定される 2 値分類結果との適合性を向上するという枠組みで、真の多値分類結果に対する事後確率最大化推定が可能であることを示す。さらに、この拡張として、任意の 2 値分類器に対して重みを持たせ、かつそれを統計的推定の枠組みでデータから決定する符号化の手続きを示す。この符号化の学習手続きによって、遺伝子発現量の実データに基く癌の多値症例分類において正解率を向上できることを示す。

2 統計的推定による2値分類器からの多値分類復号化

D 次元のデータ点 $x \in \mathcal{R}^D$ がクラスラベル $i \in C$ を持つものとする。ただし $|C| = M$ である。ラベルを指標する M 項変数 t を以下のように用意する。

$$
\begin{aligned}
t &= \{t_i\}_{i \in C} \\
t_i &= \begin{cases} 1 & \text{if } x \text{ belongs to class } i \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

誤り訂正符号方式 [4] と同様に、ラベル集合 C の重複しない任意の 2 つの部分集合について、それらを分ける 2 値分類器を考えることができる。すなわち、クラスラベルのべき集合 $2^C = \{\{1\}, \{2\}, \ldots, \{1,2\}, \ldots, C, \emptyset\}$ から、意味のない部分集合 $C, \emptyset$ を除いたべき集合 $\tilde{2}^C \equiv 2^C - \{C, \emptyset\}$ を考え、$\tilde{2}^C$ から重複のないクラスラベルの部分集合 $l, m \in \tilde{2}^C, l \cap m = \emptyset$ を選択することで、クラス集合 l 対クラス集合 m の 2 値分類問題を定義することができる。この時、$[l|m]$ をターゲットと呼ぶことにする。全ての可能なターゲットの集合を B^{AA} とする。

$$
\begin{aligned}
[l|m] \quad & \in B^{AA} = \{[1|2], [1|3], \ldots, [1|M], \ldots, [M-1|M], \\
& [12|3], [12|4], \ldots, [12|M], \cdots, [1 \ldots M-1|M], \cdots\}
\end{aligned}
$$

またその部分集合として、B^{11} すなわち $\#l = \#m = 1$ であるような $[l|m]$ の集合、あるいは B^{1R} すなわち $\#l = 1, \#m = M - 1$ であるような $[l|m]$ の集合も考えられる。ここで $\#l$ および $\#m$ はクラス集合 l および m に含まれるクラス数である。

ここで考える復号化問題は、学習データセット $\boldsymbol{L} \equiv \{(\boldsymbol{x}^{(n)}, \boldsymbol{t}^{(n)})|n = 1 : N\}$ について、ある任意のターゲット集合 B^+ に属する 2 値分類器 (の集合) が与えられた状況で、新しいデータ点 $\boldsymbol{x}$ の所属クラスを、それら 2 値分類器からの出力に基づき求めることである。データ点 $\boldsymbol{x}$ が各クラスに確率的に所属すると仮定し、$\boldsymbol{x}$ のクラス $i \in C$ への所属確率のベクトル $\boldsymbol{p}(\boldsymbol{x})$ を以下で定義する。

$$\begin{aligned} &\boldsymbol{p}(\boldsymbol{x}) \equiv \{p_i(\boldsymbol{x})\}_{i \in C} \\ &p_i(\boldsymbol{x}) \geq 0, \quad \sum_{i \in C} p_i(\boldsymbol{x}) = 1. \end{aligned} \tag{1}$$

これを用いると、$\boldsymbol{x}$ のクラス部分集合 $l \in \tilde{2}^C$ への所属確率 $p_l(\boldsymbol{x})$ は $p_l(\boldsymbol{x}) = \sum_{i \in l} p_i(\boldsymbol{x})$ で表される。

ターゲット $[l|m]$ について、学習データセット $\boldsymbol{L}$ から得られる判別関数 $f^{\boldsymbol{L}}_{[l|m]}(\boldsymbol{x}) \in \mathcal{R}$ を考える。ここで、学習アルゴリズムは何でも良く、例えば SVM であるとしておく。このとき、判別関数値 $f^{\boldsymbol{L}}_{[l|m]}(\boldsymbol{x})$ を用いて、ターゲット $[l|m]$ に関するクラス部分集合 l への所属確率を $q_{[l|m]}(\boldsymbol{x}) \equiv Pr\left(c(t) \in l | f^{\boldsymbol{L}}_{[l|m]}(\boldsymbol{x}), c(t) \in l \cup m\right)$ で表す。$c(t)$ は t の指標するクラスである。また、$q_{[l|m]}(\boldsymbol{x})$ は，ラベルに関して以下の対称性を満たすものとする。

$$q_{[l|m]}(\boldsymbol{x}) = 1 - q_{[m|l]}(\boldsymbol{x})$$

$f^{\boldsymbol{L}}_{[l|m]}(\boldsymbol{x})$ から $q_{[l|m]}(\boldsymbol{x})$ への変換には、例えば logistic regressor を用いることができる。ターゲット集合 B^+ に関するクラス所属確率をまとめて、$\boldsymbol{q}(\boldsymbol{x}) \equiv \{q_{[l|m]}(\boldsymbol{x})\}_{[l|m] \in B^+}$ で表す。$\{\boldsymbol{q}(\boldsymbol{x}^{(n)})\}_{n=1:N}$ はデータセット $\boldsymbol{L}$、2 値分類学習器、判別関数値からクラス所属確率への変換法の 3 つから決まるものであり、以下ではデータとして見なしている。

真の所属確率ベクトル $\boldsymbol{p}(\boldsymbol{x})$ が与えられたとき、ターゲット $[l|m]$ の各ラベル l, m への $\boldsymbol{x}$ の真の所属確率 $\pi_{[l|m]}(\boldsymbol{x})$ が以下で与えられるとする [2]。

$$\pi_{[l|m]}(\boldsymbol{x}) \quad = \quad \frac{p_l(\boldsymbol{x})}{p_l(\boldsymbol{x}) + p_m(\boldsymbol{x})}$$

ここでは、単一のデータ点 $\boldsymbol{x}$ の所属確率ベクトル $\boldsymbol{p}(\boldsymbol{x})$ の推定 (復号化) 問題を考えるものとして、以下では表記の簡単化のため $(\boldsymbol{x})$ を省略する。用いることのできるデータ $\boldsymbol{q}$ から $\boldsymbol{p}$ を求めたいが、両者は次元が違うため、$\boldsymbol{q}$ と同じ次元の変数 $\boldsymbol{\pi} \equiv \{\pi_{[l|m]}\}_{[l|m] \in B^+}$ を用意し、$\boldsymbol{q}, \boldsymbol{\pi}$ 間の以下の Kullback-Leibler (KL) ダイバージェンスを最小化するように、$\boldsymbol{p}$ を求める。

$$KL(\boldsymbol{q}; \boldsymbol{\pi}(\boldsymbol{p})) = \sum_{[l|m] \in B^+} \left\{ q_{[l|m]} \log \frac{q_{[l|m]}}{\pi_{[l|m]}} + (1 - q_{[l|m]}) \log \frac{1 - q_{[l|m]}}{1 - \pi_{[l|m]}} \right\} \tag{2}$$

推定の正則化のために Dirichlet 事前分布を導入して、以下の目的関数の最大化問題として定式化する。

$$\begin{aligned} V(\boldsymbol{p}) &= \sum_{[l|m]\in B^+} \left\{ q_{[l|m]} \log p_l + q_{[m|l]} \log p_m - \log(p_l + p_m) \right\} + \sum_{i\in C} \gamma_0 \log p_i + R \\ &= \sum_{k\in B^+} a_k \log p_k + \sum_{i\in C} \gamma_0 \log p_i + R \end{aligned} \tag{3}$$

ここで、γ_0 は Dirichlet 事前分布の強さを表すハイパーパラメータ、R は $\boldsymbol{q}$ のみに依存する定数である。また、a_k は次式で定義される。

$$a_k \equiv \sum_{[l|m]\in B^+, k=l} q_{[l|m]} + \sum_{[l|m]\in B^+, k=m} (1 - q_{[l|m]}) - \sum_{[l|m]\in B^+, k=l\cup m} 1 \tag{4}$$

目的関数 $V(\boldsymbol{p})$ を、式 (1) の制約の下で、$\boldsymbol{p}$ について最大化することにより、クラス所属確率の推定値 $\hat{\boldsymbol{p}}$ を得ることができる。これは、ラグランジュの未定係数法を用いた勾配法により実現できる。このように、任意のターゲット集合 B^+ について、2 値分類器の判別関数値とそれからの所属確率を用いて、クラス所属確率を推定すること、すなわち復号化ができる。以下では、この手法を MAP(maximum *a posteriori*) 法と呼ぶ。

3 2 値分類器による符号化の学習

MAP 法によれば、単一のデータ点 $\boldsymbol{x}$ のクラス所属確率 $\boldsymbol{p}(\boldsymbol{x})$ を推定することが可能である。しかし、ターゲット集合 B^+ には、2 値分類器が学習しやすいものも、しにくいものもあるにも関わらず、MAP 法では、その性質を無視して全ての分類器に一定の信頼度をおいていた。以下では、判別ロスに応じて、この信頼度を適切に設定する学習法を紹介する。

ターゲット $[l|m] \in B^+$ に対する 2 値分類器について、信頼性に基づく選択確率 $w_{[l|m]} \geq 0$ を導入し、後述するようなある目的関数を最大化するものとして、その選択確率を最適化する。すなわち、学習すべき変数はターゲット集合 B^+ に対する全ての重み $\boldsymbol{w} \equiv \{w_{[l|m]}\}_{[l|m]\in B^+}$ である。この重みは選択確率であるので、

$$w_{[l|m]} \geq 0, \quad \sum_{[l|m]\in B^+} w_{[l|m]} = 1 \tag{5}$$

という拘束を加えておく。この時、前節の KL ダイバージェンス (式 (2)) は以下のような重みつき KL ダイバージェンスとなる。

$$KL(\boldsymbol{q}; \boldsymbol{\pi}(\boldsymbol{p})) = \sum_{[l|m]\in B^+} w_{[l|m]} \left\{ q_{[l|m]} \log \frac{q_{[l|m]}}{\pi_{[l|m]}} + (1 - q_{[l|m]}) \log \frac{1 - q_{[l|m]}}{1 - \pi_{[l|m]}} \right\} \tag{6}$$

これに対応して、式 (4) は、

$$a_k \equiv \sum_{[l|m]\in B^+, k=l} w_{[l|m]} q_{[l|m]} + \sum_{[l|m]\in B^+, k=m} w_{[l|m]} (1 - q_{[l|m]}) - \sum_{[l|m]\in B^+, k=l\cup m} w_{[l|m]} \tag{7}$$

に変更される。データセット $\boldsymbol{L}$ 全体についてこれを行う必要があるので、目的関数

$$V(\{\boldsymbol{p}^{(n)}\}|\boldsymbol{w}) = \sum_{i=1}^{N} \sum_{k \in B+} a_k^{(n)} \log p_k^{(n)} + \sum_{i=1}^{N} \sum_{i \in C} \gamma_0 \log p_i^{(n)} + R \tag{8}$$

を $\{\boldsymbol{p}^{(n)}\}$ について最適化することが復号化になる。ただし、$\{\boldsymbol{p}^{(n)}\}$ の各要素は互いに独立に最適化できるので、前節の MAP 法を N 回繰り返すことで復号化可能である。また、R は $\{\boldsymbol{q}^{(n)}\}_{n=1:N}$ に依存した定数項である。

重みベクトル $\boldsymbol{w}$ は、クラス所属確率 $\boldsymbol{p}$ による判別の能力について、データセット $\boldsymbol{L}$ 全体について最適化するものとして決める。そのための効用関数 U を、クラス所属確率 $\boldsymbol{p}$ を用いた推定判別結果と真のクラスラベル $\boldsymbol{t}$ との一致度として定義する。

$$U \equiv U(\{\boldsymbol{p}^{(n)}\}, \{\boldsymbol{t}^{(n)}\}) = \sum_{n=1}^{N} \sum_{i \in C} t_i^{(n)} \mathrm{mx}(p_i^{(n)}) \tag{9}$$

ここで、$\mathrm{mx}(\boldsymbol{p}_i)$ は逆温度パラメータ β を持つ soft-max 関数

$$\mathrm{mx}(p_i) = \frac{\exp(\beta p_i)}{Z}, \quad Z = \sum_{i' \in C} \exp(\beta p_{i'})$$

であり、$\beta \to +\infty$ のとき、$\mathrm{mx}(p_i)$ は $\arg\max_i p_i$ のみ 1 とするものである。β はクラス所属確率 $\boldsymbol{p}$ を用いたクラス推定に対するノイズの大きさを制御するものであり、0 より十分に大きいものとして適当に設定する。

$\boldsymbol{w}$ の変化が $\boldsymbol{p}$ の最適化すべき目的関数 V を変えることに注意すると、ここで考える符号化学習は、学習データセット $\boldsymbol{L} \equiv \{\boldsymbol{q}^{(n)}, \boldsymbol{t}^{(n)}\}_{n=1:N}$ について、

$$\tilde{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} U(\{\tilde{\boldsymbol{p}}(\boldsymbol{w})^{(n)}\}, \{\boldsymbol{t}^{(n)}\}) \qquad \text{under the condition (1)} \tag{10}$$

$$\{\tilde{\boldsymbol{p}}^{(n)}\} = \arg\max_{\{\boldsymbol{p}^{(n)}\}} V(\{\boldsymbol{p}^{(n)}\}|\boldsymbol{w}) \qquad \text{under the condition (5)} \tag{11}$$

を満たす $\tilde{\boldsymbol{w}}$ を推定することである。データセット $\boldsymbol{L}$ 全体に対して最適化された $\tilde{\boldsymbol{w}}$ を求め、それを用いて、新しいデータ点 $\boldsymbol{x}$ のクラス所属確率の推定値 $\boldsymbol{p}(\boldsymbol{x})$ を得ることで、適切な M 値判別を行うことができる。ここで、式 (10) の最適化は、U が $\tilde{\boldsymbol{p}} \equiv \{\tilde{\boldsymbol{p}}^{(n)}\}$ を通じて間接的に $\boldsymbol{w}$ に依存しているため、陰関数微分を用いて行う必要がある。このアルゴリズムは、2 重のループ構造を成しており、内側のループで各学習データ点の復号化を、外側で重みを最適化することによる符号化を行っていることになり、以下では重み付き MAP (WMAP) 法と呼ぶことにする。

4 実験

本節では、WMAP 法を、2 種類の癌 (甲状腺癌、食道癌) [8] に関する遺伝子発現プロファイルからの癌サブクラス分類問題へ適用することで、評価を行う。評価のために、2 値分類に用いる

ターゲット集合 (1R, 11, AA)、および、2 値分類器の重み学習の有無により 6 つの識別法を準備しておく。本実験では、2 値分類器に線形カーネル SVM を、判別関数値からクラス確率への変換法として logistic regressor を用いた。MAP 法における (ハイパー) パラメータは、あらかじめ $\gamma_0 = 2, \beta = 2000$ に設定した。WMAP 法では、データに基づき符号器を学習できるため、直感的に強い 2 値分類器を必要としない。この点で boosting に類似したコンセプトである。ただし、マイクロアレイなどの遺伝子発現プロファイルは、一般的に、比較的少数次元の因子に高次元ノイズが重畳されたものとして良く近似されるため、強い (すなわち非線形性の強い) 識別器よりも比較的弱めの識別器が有効である場合が多い [6]。

ここで、2 種の癌関連遺伝子発現プロファイルの概要を以下にまとめる。

甲状腺癌発現プロファイル

本データは、168 サンプルの甲状腺がん組織に関して計測された 2000 遺伝子の発現プロファイルである。各サンプルには、臨床検査の結果に基づき濾胞腺腫 (follicular adenoma)、濾胞癌 (follicular carcinoma)、乳頭癌 (papillary adenocarcinoma)、および正常組織の 4 クラスのラベルが割り当てられており、各クラスに含まれる症例数は、それぞれ 58, 28, 42, 40 である。

食道癌発現プロファイル

本データは、日本人の食道癌組織から抽出された 1763 遺伝子の発現プロファイルである。総計 141 症例に対して、組織学的な分化型が割り当てられており、低分化型、中分化型、高分化型の 3 種に対して、それぞれ 14, 97, 30 症例である。組織分類は病理学者にも困難な課題であり、発現プロファイルを元に決めることができれば、医学的にも意義が大きい。

本実験では、さらに提案手法の性能を従来の手法と比較検討するために、多クラス SVM の実装である Weston and Watkins (WW) のアルゴリズム [7] および Crammer and Singer (CS) のアルゴリズム [3] の 2 種類を準備した。これらの手法では、提案手法と同様に線形カーネルを用いた。

各データセットに対し多クラス識別法を適用し、5-fold cross-validation (CV) を用いて性能評価を行った。結果として得られた CV accuracy を表 1 に示した。表中の括弧内の数値は CV accuracy の標準偏差である。提案手法に基く 6 種類の多クラス識別法の結果に関して比較すると、ターゲットセット B^{AA} が他のものよりも一貫して最も優れた性能をもたらしていることが分かる。甲状腺癌データセットでは，training CV accuracy が 6 種全ての多クラス識別法で上限の 1.0 に達しているため、2 値分類器の訓練に用いたデータを再び効用関数最適化に用いる WMAP 法では、判別性能を改善することが出来なかった。一方、食道癌データセットでは、training CV accuracy が上限まで達していないため、WMAP 法の training CV accuracy と test CV accuracy が MAP 法の性能と比べて改善されており、WMAP 法の有効性が分かる。また、既存の最新の多クラス識別法と比較した場合、提案した MAP、 WMAP 法は同等かそれ以上の性能を持つことが示された。

表 1: 多クラス識別法の性能比較

	MAP-1R	MAP-11	MAP-AA	WMAP-1R	WMAP-11	WMAP-AA
甲状腺癌						
Training	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
Test	0.762 (0.065)	0.762 (0.072)	**0.774** (0.074)	0.762 (0.065)	0.762 (0.072)	**0.774** (0.074)
食道癌						
Training	0.821 (0.109)	0.901 (0.0034)	0.901 (0.0034)	0.901 (0.0034)	0.917 (0.037)	0.901 (0.0034)
Test	0.695 (0.026)	0.688 (0.076)	0.696 (0.050)	0.695 (0.026)	0.688 (0.076)	**0.703** (0.051)

	WW	CS
甲状腺癌		
Training	1 (0)	1 (0)
Test	0.768 (0.069)	0.762 (0.068)
食道癌		
Training	1 (0)	1 (0)
Test	0.681 (0.056)	0.673 (0.065)

5 まとめ

複数の 2 値分類器の組合わせにより多値分類器を構成するための統計的枠組みについて紹介した。この時、多数の 2 値分類器を適切に統合することができるため、各 2 値分類器の性能は強くなくても良く、過学習を防ぐためにはむしろ弱めた方が良いと考えられる。

本稿で紹介した研究では、データに対する経験誤差最小化を行う際の目的関数として、soft-max 指数ロス関数を用いた。これは逆温度パラメータ β が無限大となる極限で、事後確率最大クラスの正解率と等価となるが、一方でこのとき事後確率の大小に関する情報は一切評価されなくなる。すなわちこのロス関数は、出力の確率解釈よりも正解率を重視していることに対応している。ただし、本稿で紹介した符号化方法は、ロス関数を微分可能なものとして種々に替えることで、各種の目的に応じて、符号表を最適化できると考えている。

本研究で用いた甲状腺癌および食道癌遺伝子発現プロファイルのデータセットは、共同研究者である大阪府立成人病センター研究所、加藤菊也所長の提供によるものである。ここに感謝申し上げる。

参考文献

[1] E. L. Allwein, R. E. Schapire, and Y. Singer, *Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers*, Journal of Machine Learning Research **1** (2001), 113–141.

[2] R. A. Bradley and M. E. Terry, *Rank Analysis of incomplete block designs I. The method of paired comparisons*, Biometrika **41** (1952), 324–345.

[3] K. Crammer and Y. Singer, *On the algorithmic implementation of multiclass kernel-based vector machines*, Journal of Machine Learning Research **2** (2001), 265–292.

[4] T. G. Dietterich and G. Bakiri, *Solving multiclass learning problems via error-correcting output codes*, Journal of Artificial Intelligence Research **2** (1995), 263–286.

[5] T. Hastie and R. Tibshirani, *Classification by pairwise coupling*, Advances in Neural Information Processing Systems (NIPS), vol. 10, 1998, pp. 507–513.

[6] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, *A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis*, Bioinformatics **21** (2005), no. 5, 631–643.

[7] J. Weston and C. Watkins, *Multi-class support vector machine*, Tech. report, University of London, 1998.

[8] N. Yukinawa, S. Oba, K. Kato, K. Taniguchi, K. Iwao-Koizumi, Y. Tamaki, S. Noguchi, and S. Ishii, *A multi-class predictor based on a probabilistic model: application to gene expression profiling-based diagnosis of thyroid tumors*, BMC Genomics **7** (2006), 190.

[9] B. Zadrozny, *Reducing multiclass to binary by coupling probability estimates*, Advances in Neural Information Processing Systems (NIPS), vol. 14, 2001, pp. 1041–1048.